人間の教示特性に基づく顔ロボットの行動学習アルゴリズム

東京理科大学　　○飯田　史也　　原　文雄　　綾井　晴美

Face Robot behavior learning based on the characteristics of human instruction

Science Univ. of Tokyo　○Fumiya IIDA　Fumio HARA　Harumi AYAI

Abstract --- We have found a difficulty in the learning of a human-friendly behavior in a face robot by means of human instruction. The difficulty is mainly spotted at the learning algorithm of the robot not taking account of characteristics of human instruction. This paper suggests an effective learning algorithm constructed by the results obtained through analysing characteristics of human instruction, and evaluates its effectiveness by computer simulation and the learning experiment using real human instruction.

Keywords: Learning-oriented Interaction, Machine learning, Face robot behavior, Characteristics of human instruction, OR-reinforcement, Comparative reinforcement

１．背景

１．１ 研究背景

近年、人間社会で活躍するロボットの期待が高まっている。中でも福祉ロボットに代表されるように、人間と直接触れ合う場でのロボットへの要求は徐々に大きくなり、人間が親和感を感じるロボットの研究も進められている[1]。著者らは人間とロボットのインタラクションを通じて、人間がどのような印象を持つかに注目し、人間が評価するロボットの印象を“ロボットの性格”と定義して実験的に分析した[2]。その結果“ロボットの性格”はインタラクション特性、すなわち人間の行動や状態に対するロボットの反応に大きな影響を受けることが解り、ロボットのsensory-motor coordinationの重要性が明らかになった。

本研究の最終的な目標は、人間と直接触れ合うロボットに人間が快く感じるロボットのsensory-motor coordinationを自己組織化することにあるが、人間の感性には大きな個人差があり、ロボットが使用される状況によっても大きく変化するため、人間が快く感じるロボットの行動を事前に決定するのは非常に難しい。さらに、ロボットに人間が快く感じる行動をさせるためには、非常に高い行動の自由度が要求され、モデルベースによるsensory-motor coordinationの構築は不可能であると考えられる。そこで本研究では学習により人間が快く感じるロボットの行動をロボット自身で獲得していくアプローチを採用し、その方法論の確立を目指す。

１．２　Learning-oriented Interaction

本研究で目指すロボットと人間の新しいインタラクションの方法論として、Learning-oriented Interactionというコンセプトを提案する。このコンセプトでは高い行動の自由度を持つ顔ロボットに人間が快く感じる行動をさせることを目指し、それによって感性情報のやり取りが必要な高度な人間とのインタラクションの実現を目標とする。感性情報を含むインタラクションには豊富な表現力に伴う高い行動の自由度が要求され、行動表現が豊富になることに起因する当事者間での行動表現における共通認識が不可欠となることが予想される。人間社会ではこのようなインタラクションは無意識に行われているが、顔ロボットと人間の間に適用するためにはロボットに学習機能を組み込む必要がある。本研究で扱う顔ロボットの行動学習の特徴としては、高い自由度の行動を学習することと、感性情報を教示データとして学習を進めることが挙げられる。感性情報には言語的な情報とジェスチャーや表情に代表される非言語的な情報が考えられるが、高い自由度の学習にはできるだけ多くの教示データが必要となることから、この両者を扱うのが妥当である。特に人間の非言語的な感性情報には非常に多くの教示データとなるべき情報が含まれており、この情報を活用することにより高い自由度の行動学習においても効率的な学習が期待できる。このように、人間から顔ロボットへの明示的な教示情報だけでなく、ロボットの学習に用いる人間の行動の中の非明示的な教示行動をNatural Instructionと呼ぶ。

１．３　問題点

Learning-oriented Interactionを実現するにあたっては、一般に以下のことが問題点として挙げられる。

①ロボットによる人間の言語的、非言語的行動の認識方法
②ロボットによる言語的、非言語的行動の実現方法
③認識した人間行動からの教示データ抽出方法
④抽出した教示データを用いた行動の獲得方法

①に関しては、画像処理、音声処理等の分野で様々な技術が開発されており、様々な人間の状態を認識する技術が開発されている。また②に関しても、人間とエージェント間のnon-verbal communicationが注目され、様々なモーダリ

ティによるインタラクションデバイスが実現されている[3]。そこで本研究では、これらの要素技術を利用して③,④の問題点に注目する。

著者らは、人間が実際に顔ロボットとのインタラクションを通じて、そのロボット行動の教示を行う際に、教示方法によっては行動の学習が行われないことがあることを指摘した[2][4]。すなわち人間による教示には特性が存在し、それが原因で従来の強化学習アルゴリズムでは学習が進まないことがある。

そこで本論文では、③のLearning-oriented Interaction中の教示データの抽出方法の議論に先立って、教示データを利用した行動の獲得方法について議論し、Natural Instructionによる行動の獲得への可能性を示す。

２． 人間の教示特性

２．１　強化学習法

強化学習法[6]は「報酬」と呼ばれる環境状態の評価情報により、エージェントの置かれる状態と実行行動の対応付けを調整する方法論であり、環境とのインタラクションにより環境から得られる「報酬」を最大化するように学習を進める。人間とのインタラクションから行動を学習する顔ロボットを扱う本研究では学習段階における人間からの教示を分析するため、追加学習が容易に行えるこの強化学習法を採用する。

強化学習法で用いられる「報酬」という概念は、通常あらかじめアルゴリズム中にエージェントが認識する環境の状態と報酬値を対応付けておくが、本研究では顔ロボットに人間の好ましい行動をさせることを目的とするため、報酬値は人間により決定されるものとする。「報酬」という概念は人間社会における学習行動と密接に関わっており、特に教育心理学の分野では様々な研究がなされている。

J.S.Bruner[5]は人間の学習における本質的特徴として、
①検証手続きの定式化、
②検証手続きの実施、
③検証結果と何らかの基準の比較
という手順があることを指摘し、③が行われるときには修正の知識が必要であり、更に目的の達成に向かっているかどうかという指針があることが望ましく、教授者はこれらを学習者に示すことが望ましいと述べている。これは教授者から学習者への報酬に他ならず、人間にとって直感的にわかりやすい概念であると考えられ、本論文で提案するNatual Instructionに応用することのできる人間の教示行動としての一般性をもつと考えられる。

２．２人間の教示特性

著者らはこれらのコンセプトに基づき、実際に人間の教示に基づく顔ロボットの行動学習の実験を行ったが、非常

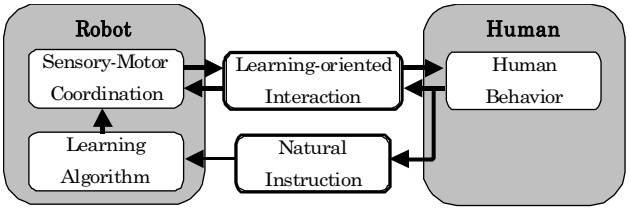

図１ Learning-oriented Interaction概念図

に低い自由度の行動学習にも関わらず、被験者の教示法によって学習が進まないことがあることを指摘した[2]。この原因は、このロボットの学習アルゴリズムに人間の教示特性が考慮されていないことにあることが明らかになった。すなわち、顔ロボットへの人間の教示には目的の達成に向かっているかどうかの指針が含まれており、人間が顔ロボットにある特定の行動を実行させることを望んだ場合、人間はそのロボットの行動目標の達成が部分的であってもそれが達成に向かっていれば正の「報酬」を与えるが、通常の強化学習アルゴリズムを備えたロボットは正解行動への道筋を示す情報が入っている報酬を活用できずに学習が進まなくなる。

そこで本論文ではこの人間の教示特性に基づいた行動学習アルゴリズムを提案し、実際の人間の教示による行動学習実験でその有効性を示す。

３． 行動学習アルゴリズム

３．１　基本設定

これらの問題を考えるにあたって、以下のように問題設定を行う。

①ロボットには複数のアクチュエータが備えられている。このアクチュエータの数を行動の自由度と呼ぶ。

②ロボットがある状態に対して行う行動は各アクチュエータの出力値の論理積の形で表現できるものとする。

③人間の教示はスカラー値で、ロボットが１回の行動を実行する毎に与えられるものとする。

強化学習の土台となるＱ値の更新式は

$$Q(s,a) := (1-\alpha)Q(s,a) + \alpha r \qquad (1)$$

とする。ここでｓはロボットが認識した環境の状態、ａは実行した行動、αは学習率、ｒは人間の教示による報酬値とする。一回の行動に対する強化でどの行動に対応するＱ値を上式で更新するかを議論する。また状態ｓの数が増えれば学習回数が増えることになるが、ここでは議論の対象とせず、１つの状態に対する学習を取り扱うものとし、以降、Q(s,a)をQ(a)と記す。

３．２　行動学習アルゴリズム

上記の問題設定のもとで人間の教示特性を考慮に入れた学習アルゴリズムとして、「AND強化型学習アルゴリズム」を

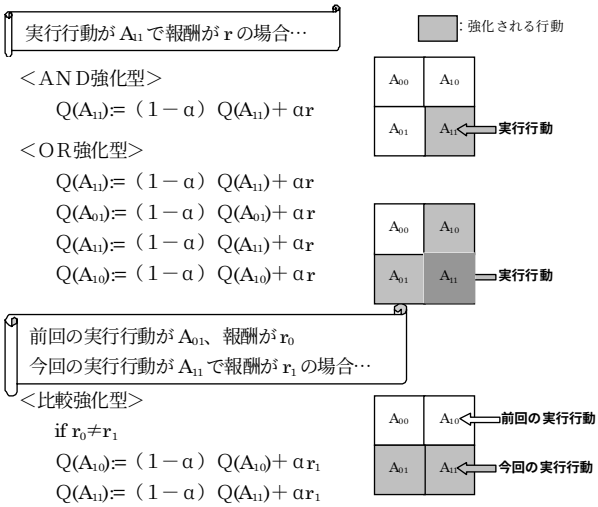

図2 各学習アルゴリズムの強化方法
（行動自由度２の場合）

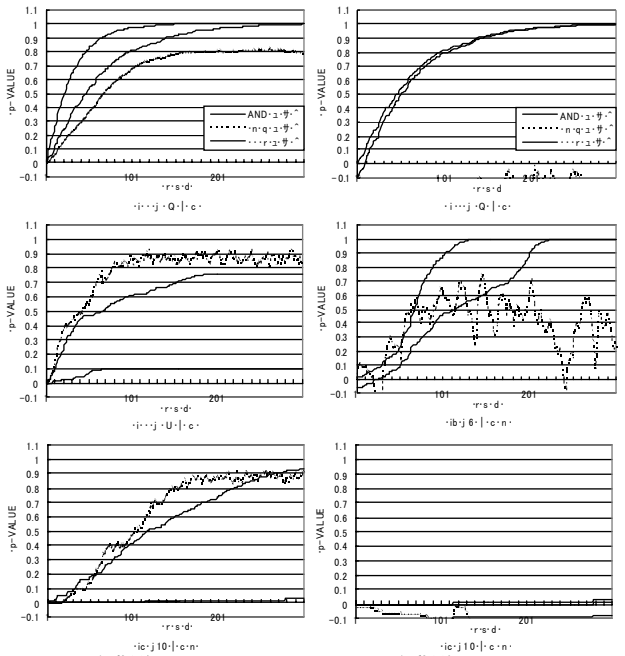

図3 各学習アルゴリズムによるQ値の推移

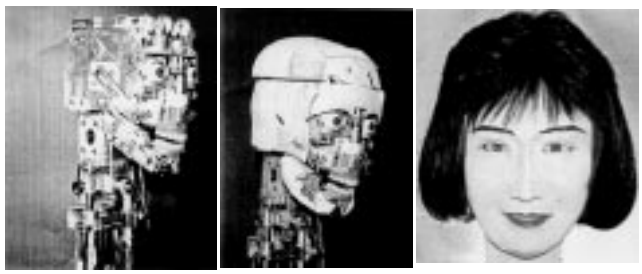

図4 顔ロボット概観

比較の基準アルゴリズムとし、「OR 強化型学習アルゴリズム」および「比較強化型学習アルゴリズム」を提案する。以下に各学習アルゴリズムの内容を示す（図１参照)。

＜ AND 強化型＞

この学習アルゴリズムは以下に提案する２つのアルゴリズムとの比較のために用意する。従来の強化学習と同様、与えられた「報酬」を実行した要素行動の論理積で表される行動に対応するＱ値のみの強化に使う。

＜ OR 強化型＞

与えられた「報酬」を実行した行動に対応するＱ値だけでなく、実行行動の行動要素を含む行動に対応するＱ値も同様に強化を行う。ただし、実行行動に対応するＱ値は行動の自由度数分だけ強化される。

＜比較強化型＞

人間が与える「報酬」が前回の結果との比較評価に基づいているという教示特性を利用して、同じ状態の 1STEP 前の実行行動および「報酬」を記憶しておき、今回の実行行動および報酬と比較して強化を行う。前回と比較して「報酬」が変化した場合、前回の行動と今回の行動で変化した要素行動を含む行動集合を強化する。「報酬」が前回と変わらなかったときは今回実行した行動集合のみ強化する。この強化型の特徴はOR強化型と違って正解行動を負強化することがない。

４． シミュレーション

４．１ シミュレーション方法

提案した学習アルゴリズムの評価をシミュレーションを用いて行う。シミュレーションでは、1STEP 毎にロボットが出力する行動に対して、決められた正解行動と比較した行動の評価を「報酬」として顔ロボットに与え、この「報酬」を基に提案したそれぞれの学習アルゴリズムで行動を学習して行く。本実験では各アクチュエータの出力値は 2 種類（ON,OFF)、人間が顔ロボットに与えることが出きる「報酬」は○または×の２種類（それぞれ報酬値 +1.0，-1.0）であるとする。[2]での実験に基づいて、顔ロボットの行動に対する人間の「報酬」の決定則として以下の２つを用意する。

①行動評価モデル A

同じ状態の前回の行動と比較して、満足できる行動要素の数が増加した場合は○報酬を与え、減少した場合は×報酬を与える。同じ場合は○報酬を与える。ただし、すべての行動要素が満足できる場合には○報酬とし、逆にすべての行動要素が満足できない場合には×報酬とする。

②行動評価モデル B

行動目標がすべて達成されたときのみ○報酬を与え、それ以外は×報酬を与える。

学習率α＝0.1、ボルツマン温度＝0.1、行動自由度＝2，6，10 のもとでシミュレーションを行った。

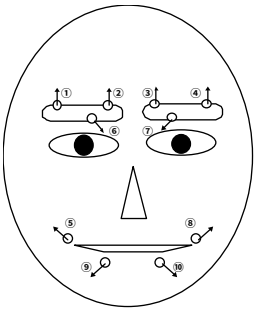

図５ ACDIS番号と表情制御点の対応

表１．○報酬を与えたときの各制御点の出力値

	出力値	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
被験者１	1	77%	27%	48%	78%	90%	36%	55%	74%	22%	32%
	0	23%	73%	52%	22%	10%	64%	45%	26%	78%	68%
被験者２	1	61%	74	26%	58%	97%	38%	22%	72%	39%	61%
	0	39%	26%	74%	42%	3%	62%	78%	28%	61%	39%
被験者３	1	22%	18%	25%	59%	98%	55%	66%	95%	50%	36%
	0	22%	82%	75%	41%	2%	45%	34%	5%	50%	64%
被験者４	1	73%	38%	41%	72%	86%	39%	61%	87%	67%	43%
	0	27%	62%	59%	28%	14%	61%	39%	13%	33%	57%
被験者５	1	51%	47%	51%	59%	71%	51%	49%	94%	59%	53%
	0	49%	53%	49%	41%	29%	49%	51%	6%	41%	47%

表２．各被験者の正解表情（幸福） ＊：非主成分

	ACDIS番号 ①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
被験者１	･P	･O	･･	･P	･P	･･	･･	･P	･O	･･
被験者２	･･	･P	･O	･･	･P	･･	･O	･P	･･	･･
被験者３	･･	･O	･O	･･	･P	･･	･･	･P	･･	･･
被験者４	･P	･･	･･	･P	･P	･･	･･	･P	･･	･･
被験者５	･･	･･	･･	･･	･P	･･	･･	･P	･･	･･

表３．正解表情・不正解表情に対する報酬分布

	･･･ｱ･ﾒ･P	･･･ｱ･ﾒ･Q	･･･ｱ･ﾒ･R	･･･ｱ･ﾒ･S	･･･ｱ･ﾒ･T
･ｳ･･･ﾌ･････V	90%	98%	94%	93%	99%
･s･ｳ･･･ﾌ･~･･･V	25%	14%	36%	5%	19%

表４．正解表情に近づいたときの報酬分布

	･･･ｱ･ﾒ･P	･･･ｱ･ﾒ･Q	･･･ｱ･ﾒ･R	･･･ｱ･ﾒ･S	･･･ｱ･ﾒ･T
･････V	72.2%	27.3%	54.3%	81.5%	56.3%
･｢･･･V	19.0%	52.3%	30.4%	15.4%	39.6%
･~･･･V	8.9%	20.5%	15.2%	3.1%	4.2%

表５．正解表情から離れたときの報酬分布

	･･･ｱ･ﾒ･P	･･･ｱ･ﾒ･Q	･･･ｱ･ﾒ･R	･･･ｱ･ﾒ･S	･･･ｱ･ﾒ･T
･････V	13.3%	13.1%	5.8%	21.7%	11.8%
･｢･･･V	53.3%	60.7%	25.0%	66.7%	66.7%
･~･･･V	33.3%	26.2%	69.2%	11.6%	21.6%

４．２　結果および考察

図3は行動評価モデルAと行動評価モデルBに対して、学習ステップ毎の正解行動に対応するQ値の推移を、行動の自由度毎に示している。すなわちこのQ値が高くなるほど、正解行動の学習が進んだことを示す。それぞれのQ値は10回の試行で得られた結果を平均した値である。行動評価モデルAについては自由度が上がるにつれて、従来のAND強化型と比較して、OR強化型と比較強化型の学習アルゴリズムの効果が顕著に現れていることが解る。行動評価モデルAに対してはOR強化型は全体的に比較強化型よりも学習が速く収束しているが、行動評価モデルBに対する学習では正解行動を負強化してしまう性質により安定した強化が行われず、自由度が高くなった場合は正解行動を負強化する行動の方が選択確率が高いため結果的に正解行動を負強化してしまう傾向にあるために振動的な学習が行われる。

５．　実験

５．１　実験目的

本実験は実際に人間と顔ロボットのインタラクションの中で人間による教示が行われた場合に、提案した学習アルゴリズムが有効であるかどうかを評価することを目的とする。各学習アルゴリズムはシミュレーション実験により行動の自由度が大きくなった場合に顕著に現れるという知見が得られたので、本実験では行動の自由度が大きい、顔ロボットによる表情の学習を問題として扱う。

顔ロボットは、人間と表情等のnon-verbal communicationを含んだインタラクションを行うことを目的として開発されたロボットで、人間の頭部動作を模倣できる(図4)。構造は大きく分けてフレーム部、頭蓋部、皮膚部からなり、フレーム部には最大18本のACDISと呼ばれる空気圧アクチュエータを備え、皮膚と接合し、伸縮動作をさせることでさまざまな表情を表出することができる。さらに首と眼球にもモータ等のアクチュエータを備え、全部で24自由度を持つ。本論文ではシミュレーションで検証した10自由度までを扱うために、表情の表出に強く影響する10自由度分のACDISを選び、それらを学習対象の行動要素とする。選んだACDISとその制御点の対応を図5に示す。

５．２　実験方法

被験者には顔ロボットが一つの表情を学習するプロセスで一連の教示を行ってもらい、学習の収束性を評価基準に学習アルゴリズムの有効性を議論する。

実験は「AND強化型」、「OR強化型」、「比較強化型」の3つに対して行い、それぞれの学習パラメータは被験者の負担を小さくするため、シミュレーション実験よりも早く学習が収束するように学習率α＝0.3、ボルツマン温度T＝0.1とした。

被験者には予め目標の表情を伝えておき、毎ステップ表出する表情が目標の表情にみえるかどうかを○、△、×で評価してもらう（報酬値はそれぞれ+1.0, 0.0, -1.0)。評価を3段階にすることで被験者が相対評価と絶対評価を行うことができる。目標表情は、表情としての特徴が大きい「幸福」の表情に対して行う。

５．３　実験結果および考察

（1）　評価方法

人間の教示特性を調べるために、まず各被験者がどの表情

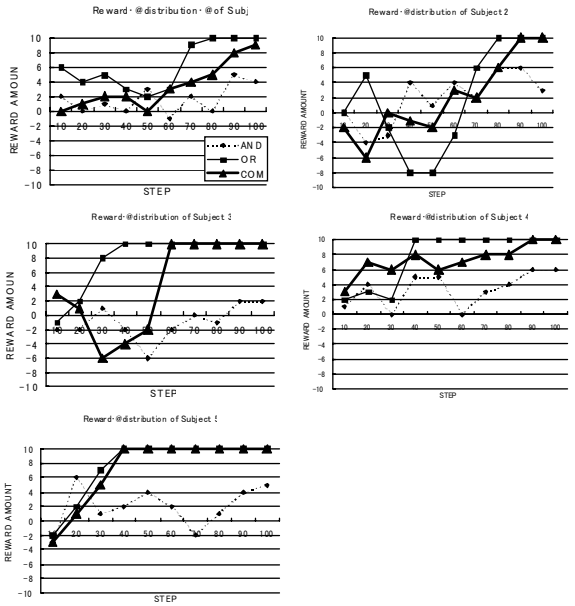

図６ 各被験者の報酬分布推移

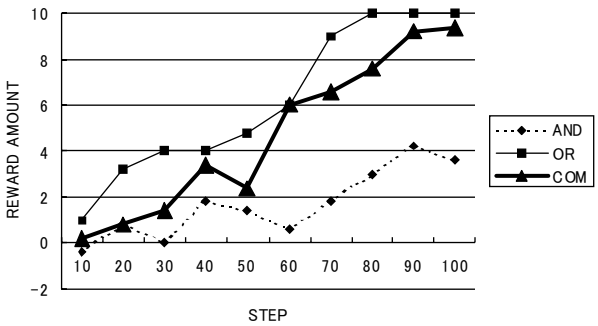

図７平均報酬分布推移

表６ 平均学習効率

STEP	10	20	30	40	50	60	70	80	90	100	average
AND	-0.4	0.8	0	1.8	1.4	0.6	1	2	4.6	4	
OR	1	3.2	4	5	4.8	6	9	10	10	10	
COM	0.2	0.8	1.4	3	2.4	6.6	6.8	7.8	9.6	9.8	
(OR-AND)/10	14.0%	24.0%	40.0%	32.0%	34.0%	54.0%	80.0%	80.0%	54.0%	60.0%	47.2%
(COM-AND)/10	6.0%	0.0%	14.0%	12.0%	10.0%	60.0%	58.0%	58.0%	50.0%	58.0%	32.6%

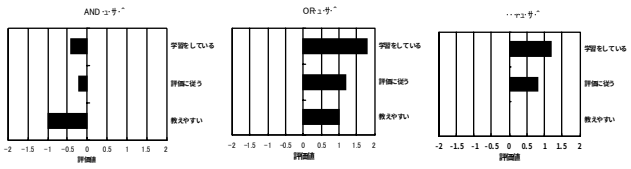

図８ アンケート結果

を幸福の表情としているかを調べる。本実験では表情を定義する基準を、被験者が与えた報酬分布から推測するしかないので、まず全実験を通じて各被験者が○報酬を与えたときの各 ACDIS の出力値を調べ、それを基に各被験者の正解表情を定める。本実験では、各学習ステップごとに ACDIS の出力値は 0（縮）、1（伸）と決めたので、各 ACDIS が 0 を出力した回数と 1 を出力した回数を、全学習ステップ通じて数えることができる。そこで、各被験者が○報酬を与えた場合のみの 0 と 1 の出力回数を求め、それを表 1 にパーセント表示で示す。

表 1 から、各被験者が○報酬を与えた場合の ACDIS の出力値には偏りがあることがわかる。この表は学習過程における ACDIS の出力値を基に計算されているが、実験中に学習が収束した場合の出力値もカウントされていることと、各被験者に対し 300 ステップの教示を行ってもらっていることを考慮に入れると、被験者が正解と評価する表情の偏りをあらわしていると考えられる。

そこで、表 1 を基に経験的に決めた 70%の閾値を超えた ACDIS の出力値を正解表情と定義し、超えなかったものに対してはその被験者にとっての幸福の表情に関して非主成分であるとする。表 2 に各被験者の正解表情を示す。

（2） 人間の教示特性の検証

提案する行動学習アルゴリズムの有効性を分析するために、まず本論文で指摘する人間の教示特性の検証を行う。

最初に 4.1 で定義した人間の行動評価モデルが実際の人間の教示の中にどの程度あるかを調べる。表 3 は顔ロボットが正解表情を表出した際に○報酬を与えた割合と、不正解表情を表出した際に×報酬を与えた割合を示し、各被験者がどれだけ行動評価モデルの A または B に近いかを定量的に知ることができる。すなわち、正解表情を出力したときに○報酬を与え、非正解表情を出力したときに×報酬を与える割合が高いほど、行動評価モデル B（絶対評価型）に近いといえる。本実験では自由度の高い問題を扱っているため、全体として行動評価モデル A（相対評価型）に近いものと考えられる。

次に表情の出力が正解の表情に近づいたときの報酬値の分布を調べる。ここで正解の表情に近づいたかどうかを客観的に測るための基準を以下のように定める。すなわち、それぞれの被験者に対して決めることができた正解表情における各 ACDIS の出力値と各ステップにおける各 ACDIS の出力値を比較し、一致している出力値の個数を正解表情からの距離と定義する。本実験では 10 個の ACDIS を用いたことから、正解表情を出力したときの距離は 0、最大距離は 10 となる。

次に 1STEP 前の距離と今回の距離を比較し、距離が近くなっていたときの報酬値の分布を調べる。表 4 は距離が近くなった場合に○、△、×の報酬を与えた回数を距離が近づいた回数全体でで割った割合である。表 5 は同様に距離が離れた場合の計算結果である。

表 4 と表 5 を比較すると、各被験者とも正解表情に近づいた場合は明らかに○報酬を与える割合が多く、逆に離れた場合には×報酬を与える割合が多い。これは人間の教示に行動評価モデル A（比較行動評価型）の特性があることを示している。

（3） アルゴリズムの有効性

以上より、人間の教示の中には行動評価モデルAの特性が含まれることが確認できたので、これらの特性が本研究で提案する行動学習アルゴリズムに対して有効であるかどうかを検討する。図6は各アルゴリズムに対する100STEPの学習過程で与えられた報酬値を10STEP毎に累積したものである。いずれの被験者に対しても、従来のAND強化型よりも高い報酬値を得ており、有効性が確認できる。特に行動評価モデルAの特性が強い、被験者3，4，5についてはAND強化型と大きな差が確認できる。被験者1については正解表情の行動自由度が高いために最初はあまり差が現れないが、徐々に差が広がっているのがわかる。

図7には各被験者の報酬累積値を学習アルゴリズムごとに平均したものを示す。全体としてOR強化型、比較強化型は、AND強化型に比較して高い報酬値を得ている。シミュレーションと違って、OR強化型の方が比較強化型よりも高い値となっているのは、正解行動に近づいたとき、または離れたときに与えられる報酬値の非一貫性が、比較強化型には外乱となって学習を妨げていると考えられる。

また、表6は平均学習効率を計算したものである。従来のAND強化型に比べて全体を通して効率的な学習が行われており、特に後半に差が広がっているのはAND強化型で収束できない教示特性の被験者による差が顕著に表れたと考えることができる。また、各強化型に対する行動教示実験が終わった直後に被験者に以下の質問に対する回答を得た。

①学習をしている - 学習をしていない
②評価に従うか - 反抗的か
③教えやすいか - 教えにくいか

アンケートはこれらの問いに対して５段階の評価をしてもらい、図8は5段階の評価をそれぞれ+2，+1，0，-1，-2点に換算して５人の被験者に対して平均した結果である。これらの結果は図7に示した報酬の推移の結果とよく対応しており、教示者である人間の心理的にもOR強化型、比較強化型の行動学習アルゴリズムが有効であることを示している。

６． 結論および今後の展望

本論文では顔ロボットが人間とのより親密なインタラクションを実現するための基本コンセプトとなるLearning-oriented Interactionを提案した。さらにLearning-oriented Interactionの必要不可欠な要素となるNatural Instructionによる顔ロボットの行動学習アルゴリズムには人間の教示特性を考慮に入れないと顔ロボットは行動を学習できないことがある事実に基づき、人間の教示特性を定量的に分析するとともに、その教示特性を考慮した学習アルゴリズムを提案した。またこれらのアルゴリズム有効性ををシミュレーションおよび実際の人間とのインタラクションから検証した。

以上は顔ロボットを対象に検証された結論であるが、人間に優しいロボットの実現において人間とロボットとのインタラクションは大切であると考えられることから、次のことがいえる。人間の教示に基づいてそのようなロボットが行動の学習を進める場合には、人間の教示方法の一般的特性を考慮に入れることで､行動の評価情報が少なくても、高い自由度の行動を学習できる可能性がある。そして更なる人間の教示特性を学習アルゴリズムに適用することが望まれる。

また､本論文で提案した学習アルゴリズムを実際にNatural Instructionに応用するにあたり、人間の行動情報をどのように「報酬」に割り当てて行くのかが今後の課題であり、この問題点を解決することによりLearning-oriented Interactionが実現できると考えられる。

[参考文献]

[1]中田，佐藤，森，溝口：“ロボットの対人行動による親和感の演出”，日本ロボット学会誌，ｖｏｌ．１５，no.7,pp1068-1074,1997

[2]F. Iida, M. Tabata, F. Hara: Generating Personality Character in a Face Robot through Interaction with Human, Proc. of 7th IEEE International Workshop on Robot and Human Communication, pp481-486, 1998

[3]F.Hara, H.Kobayashi: “State-of-the Art in Component Technology for an Animated Facerobot- Its Component Technology Development for Interactive Communication with Humans”,The Intl. Jour. of the Robotics Society of Japan Advanced Robotics, Vol.11, No.6, pp585-604, 1997

[4]飯田史也、原文雄：“人間の教示特性に基づくロボット行動学習アルゴリズムの提案”,第16回日本ロボット学会学術講演会予稿集，pp655-656，1998

[5] J.S.ブルーナー：教授理論の建設，pp.74 - 78，黎明書房，1983

[6]浅田稔：“強化学習の実ロボットへの応用とその課題”，人工知能学会誌，Vol.12，No.6，pp831-836，1997